An Image Augmentation Method Using Convolutional Network for Thyroid Nodule Classification by Transfer Learning

Ye Zhu
SJTU-ParisTech Elite Institute of Technology
Shanghai Jiao Tong University
Shanghai, China
e-mail: Lsesasfs@sjtu.edu.cn

Zhuang Fu
State Key Lab of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, China
e-mail: zhfu@sjtu.edu.cn

Jian Fei
Ruijin Hospital
Shanghai Jiao Tong University
Shanghai, China
e-mail: feijian@hotmail.com

***Abstract*—Thyroid nodules are severely jeopardizing our health. In the diagnosis of thyroid nodules, the ultrasound images serve as an essential tool to discriminate the malignant nodules from the benign ones. In this paper, the method of transfer learning is applied to classify the malignant and benign thyroid nodules based on their ultrasound images. The principal steps are preprocessing, data augmentation and classification by transfer learning. The preprocessing concentrates in extracting the region of interest (ROI). Two techniques of data augmentation are realized in our experiment, the traditional ways of augmenting images and a small convolutional network proposed by our own. After the augmentation of dataset, a pre-trained residual network is adopted to do transfer learning, and the parameters of this pre-trained net are fine-tuned with three different datasets that we have attained, including the original dataset, the augmented dataset via traditional methods and the augmented dataset via our convolutional network. The performances are then evaluated by three indexes, and the final results have proved the effectiveness of our convolutional augmentation network as well as the application of transfer learning. The best accuracy on the augmented dataset via convolutional network attains 93.75%, which exceeds the results of other two datasets and in the meanwhile outperforms other relevant methods.**

Keywords-image augmentation; convolutional network; thyroid nodules; classification; transfer learning

I. INTRODUCTION

In this paper, we propose an image augmentation method based on convolutional network and then evaluate its effectiveness by comparing the final classification results with the dataset augmented by traditional methods. The final classification is realized by transfer learning to fine-tune a pre-trained ResNet-18. The results have proved that our convolutional network for augmentation has a better performance than the traditional methods and is capable of generating a larger dataset from the initial restrained dataset.

As one of the most common frequent endocrine carcinoma diseases, thyroid nodules are severely jeopardizing people's health. Nowadays, with more thyroid nodules being detected, it is urgent to distinguish the possible malignant nodules from the benign ones. The existing diagnosis methods include sonography for primary diagnosis and fine needle aspiration (FNA) or surgeries for biopsies. Nevertheless, the primary diagnosis by sonography relies, to a great extent, on the experience and judgments of radiologists, which may increase the misdiagnosis rate and a reliable computer aided diagnosis (CAD) system has become particularly significant under these circumstances [1]. A fully-automatic CAD system has been the pursuit of many researchers in the field of medical image processing [2]. In the work of Ginneken [3], CAD is developed for detection of lung nodules. Zekeri et al. proposed a method to extract effective texture features for diagnose different breast nodules [4]. A CAD system using extreme learning machine is proposed by Li et al. in [5]. Liu et al. have also adopted the method of transfer learning for thyroid nodules classification, however, instead of classifying directly thyroid nodules by deep CNNs models, they have used it as a feature extractor, and combined the features extracted by deep learning model with lower features extracted by traditional machine learning methods [6]. Luo et al. tried to classify the thyroid nodules via linear discriminant analysis [7]. There are also some researchers who use textural and shape features as basis for classification, such as Legakis et al. [8]. Multiple-instance learning has also been tested by Ding et al. for thyroid ultrasound image classification [9] based on local texture analysis [10].

Deep learning models such as CNNs has proved its efficiency in various learning tasks, including the image classification problems. However, training a deep convolutional neutron network from beginning requires enormous number of images while the medical images are usually more difficult to gather and more cumbersome to process due to their particularities. Lack of sufficient images will result in problems like over-fitting, thus two possible solutions are transfer learning and data augmentation. Transfer learning adopts pre-trained deep learning models and then fine-tuning the parameter with existing images in purpose of adjusting the pre-trained model to fit the current classification problem. As for data augmentation, the classical methods for augmenting images such as cropping, rotation, flipping and rescaling. But unlike other images that can easily be labeled and recognized, medical images request well trained physicians to classifier various type of diseases. Additionally, traditional way of augmenting image data risk eliminating the paramount region of the image by random cropping, such as the tumor in an ultrasound Images. As a consequence, an efficient augmentation method for medical images seems to be of great value for a more puissant CAD system and more precis diagnosis for patients. The convolutional layers in various deep learning models play the role of feature extractor. Inspired by the work of [11], we make an effort to build a small convolutional network to generate new image data based on the original ones.

II. Dataset and Preprocessing

The dataset we use for experiments comes from an open access provided by Universidad Nacional de Colombia. Instituto de Diagn ostico Medico [12], which contains in total 298 valid cases (several cases contain more than one ultrasound image), and the images are in form of RGB. Among the 298 cases with a TI-RADS score, 246 cases are diagnosed as malignant (has a TI-RADS score 4a, 4b, 4c or 5), while 52 cases are benign with a TI-RADS score 2 or 3. Thyroid Imaging Reporting and Data System (TI-RADS) scores are a reference and potential indication for the character of the thyroid nodules detected, they are specified as 2, 3, 4a, 4b, 4c and 5, which can be interpreted respectively as “not suspicious”, “probably benign”, “possess one suspicious feature”, “possess two suspicious features”, “possess three or more suspicious features” and “probably malignancy”. And the suspicious features mentioned above represent some suggestive features in sonography, exemplified by micro-calcifications, hypo-echogenicity, absence of a halo, intra-nodular flow, solidity and taller-than-wide shape [13].

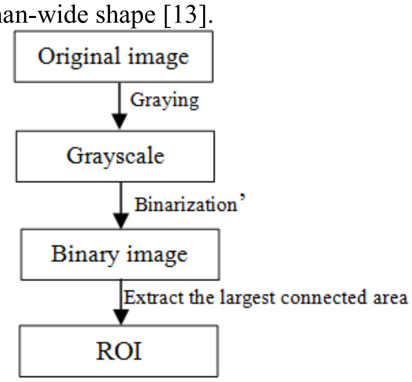


Figure 1. Algorithm for ROI extraction.

The preprocessing is mainly about extracting the region of interest (ROI) since the original ultrasound images contains an artificial background which is not appropriate for deep learning. The algorithm we use for ROI extraction can be described as Fig. 1. Fig. 2 is an example of ROI.

We can observe from the original ultrasound image that the ROI is actually represented by the largest central area. The principle of this algorithm is to locate the largest conntected area in the binary imgae and then extract the corresponding region in the original RGB image.

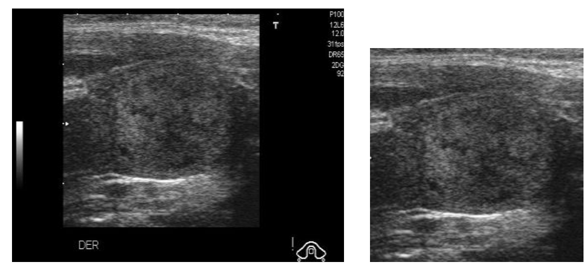


Figure 2. A example of an original image and the ROI extracted.

III. Data Augmentation

A. Tranditional Transformations

Traditional methods for augment image data include some classical transformations, such as cropping, flip, rotation, shift, rescale and contrast variation. For each input image, the classical transformation will generate a new image using the above transformation. In our experience, we have majorly adopted random cropping, random flip and rescale as means of augmentation.

B. Convolutional Network

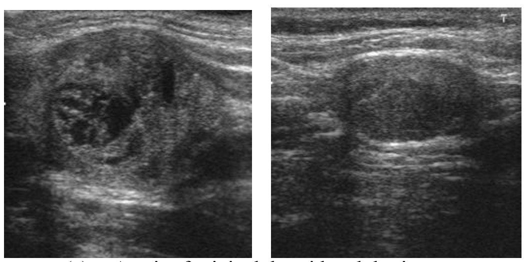
In classical deep learning models as CNNs, the convolution layers are an indispensable part of the overall configuration, and they as a feature extractor for image data. Unlike the traditional ways that risk to cut off the pivotal region of medical images, using the convolutional network to generate new images enable us to well maintain the essential feature information like the nodules and the surrounding tissues. In this section, we propose a “small CNN” that contains only convolutional layers to realize the image data augmentation, and we refer it as “Augmentation Net”. The configuration of this augmentation net is given as follows.

1. Conv with 6 channels and 3×3 kernel. Relu activations.
2. Conv with 16 channels and 3×3 kernel. Relu activations.
3. Conv with 3 channels and 3×3 kernel.

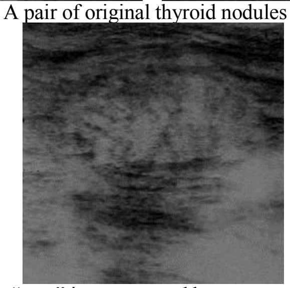
The input of this network are two ultrasound images of thyroid nodules in a same class, i.e. two malignant nodules or two benign nodules, and the output of this network is a new image of the same class that has mixed the major features of two input images. The input layer is 6 channels deep for two RGB images, and the output is 3 channels deep which represents a single RGB image generated in a same image size. By using this net, the number of newly created images is N^2 with N original images. Additionally, we have computed a content loss for the augmentation net in purpose of regularizing the output image. The content loss is defined as the (1), which is the mean square error between the augmented image A and the target image T, divided by the total area of the image, in which *l* is the length of image while *w* is its width.

$$L_{content} = \frac{1}{lw}\sum_{i,j} A_{i,j} - T_{i,j} \tag{1}$$

Fig. 3 is an example of the augmented image generated by the augmentation net above. We can observe from the given example that the new image possesses several visible features that are similar to the original ones, but has a distinguish difference in terms of contrasts, colors and some other image features, which signifies primarily that the augmented image is a "new" image.



(a) A pair of original thyroid nodules images



(b) A "new" image created by augmentation net

Figure 3. A example of the image augmentation

The dataset is then divided into training and validation sets as Table 1 presents. There are three datasets in total: the original dataset without image augmentation, the dataset with new images augmented by traditional methods and the dataset with our augmentation net. We refer them respectively as "Original", "Experiment 1" and "Experiment 2" in the rest of the paper for comparisons.

TABLE I. DISTRIBUTIONS OF DATASETS

Datasets	Train		Validation	
	Malignant	*Benign*	*Malignant*	*Benign*
Original	222	46	60	15
Experiment 1	600	120	200	40
Experiment 2	600	120	200	40

IV. EXPERIMENTS BY TRANSFER LEARNING

The experiments are conducted by transfer learning. The pre-trained deep learning model we have adopted is a ResNet18 in Pytorch, we then fine-tune the parameters with the image dataset that has been divided into training set and validation set.

A. *Residual Net*

Using deep residual net for image recognition is first brought up by K. He et al. [14] in the year of 2015, and it is the network that has become the champion of ImageNet Large Scale Visual Recognition Competition 2015 (ILSVRC).

In the traditional deep learning field, we are often perplexed by the problem of degradation. When the depth of network keeps increasing, the accuracy will saturate and then begin to decrease. This degradation in accuracy is caused by training error for deep models, and the residual net is proposed to solve this problem. Instead of fitting each stacked layer directly to a desired underlying mapping as H(x), the residual net allows these layers to fit a residual mapping F(x) := H(x) – x. By this method of residual learning, the underlying mapping H(x) is realized by some feedforward neural networks with certain shortcut connections [14] as shown in Fig. 4. These shortcuts enable us to train the model more rapidly and more accurately without adding extra parameters to the network

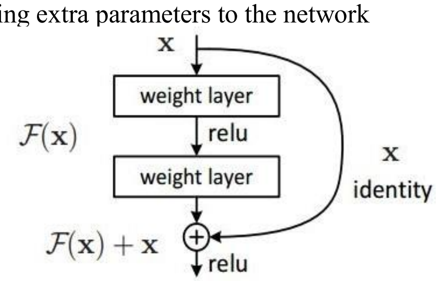


Figure 4. A building block for residual learning from [14]

In terms of medical images like thyroid nodules, it is worth noting that medical images are not "natural", while the existing deep learning models are all trained with natural images. As a consequence, the problem of overfitting and degradation may be even more severe when it comes to medical images. The choice of residual net as our experimental method has taken its pertinence to degradation into considerations. The variation of residual net includes ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152, the model we have adopted is the 18-layer ResNet-18 and modified its final classification numbers to two classes.

B. *Transfer Learning*

Transfer learning is a method that uses the pre-trained models to fit the new problems by usually fine-tuning the parameters, which enables us to avoid burdensome computations for "training from zero" The principle of transfer learning is the inner connections among different learning tasks, just as human being's learning behaviors, we can always unearth some connections between new knowledge and the previous knowledge that we have already acquired.

We have conducted three experiments based on three datasets, for each dataset, we have trained the pre-trained ResNet-18 for 30 epochs, with an exponential learning rate decay of factor 0.1 for every seven epochs. The loss function for this transfer learning step is cross entropy.

All the experiments including image augmentation and transfer learning are realized in Pytorch, on GPU, the training time for the original dataset and the augmented dataset are 3m16s and 8m36s.

V. EXPERIMENTAL RESULTATS

The experimental results are presented in this section to demonstrate the effectiveness of the augmentation net in generating new ultrasound images for thyroid nodules.

Three experiments are evaluated by three indexes: accuracy, sensitivity and specificity. Accuracy is computed by *(TP + TN)/ (TP + TN + FP + FN)*, sensitivity equals *TP /(TP+FN)* and specificity is *TN/(TN+FP)*.

TABLE II. COMPARISON RESULTS OF DIFFERENT EXPERIMENTAL SETS

Experiments	Accuracy	Sensitivity	Specificity
Origial	84%	84.48%	82.35%
Experiment 1	92.08%	92.96%	87.80%
Experiment 2	93.75%	93.96%	92.68%

We can observe from the above table that both augmentation methods help markedly improve the classification results, which also prove the fact that image data augmentation is an effective method in image classification problems especially when the dataset is restricted. The classification accuracy increases from 84% to more than 92% after augmentation of data image.

As for the comparison between two augmentation methods, the experimental results have shown that it is our proposed augmentation net that has a better performance than the traditional methods, with an exceeding accuracy of more than 1%. The same preponderances also appear in other two indexes.

The accuracy plot at each epoch is shown in Fig. 5, the total number of training epochs is 30. The figure reveals the same conclusion as we have mentioned, the dataset we obtained with our augmentation net has a slightly better performance than the dataset obtained by traditional methods. And the overall plot has also verified the fact that the image data augmentation helps in preventing and reducing the phenomenon of overfitting. The phenomenon of convergence is also observed in this plot.

Compared with other methods to classify the malignant and benign thyroid nodules, our method achieves a better performance that attains the classification accuracy of 93.75%. For example, the best classification accuracy in [6] is 93.1%, [7] has an accuracy of 87.5%, and in [8], the corresponding number is 93.2%.

In the meanwhile, we need to pay close attention to the configuration of our augmentation net. The final one we have shown in the paper is three-layer structure, this number of layer proposed has taken the characteristics of medical ultrasound images into account. Due to the facts that the medical ultrasound images usually have a lower contrast and lower luminosity than common natural images, and the features shown are more univocal, therefore, we do not recommend this convolutional net for augmentation exceeds more layers than three, which will produce unrecognizable images. Generally speaking, according to our experiments, a two-layer or three-layer convolutional network is recommended.

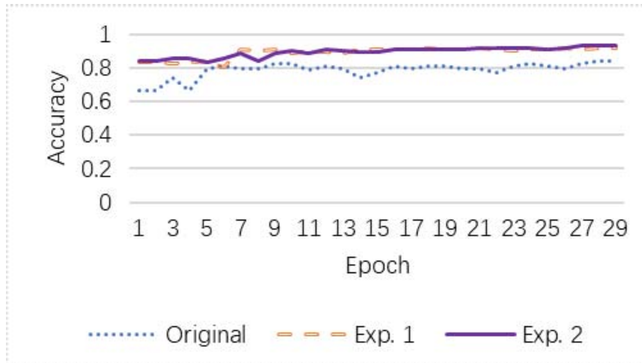


Figure 5. Accuracy plot for different datasets.

VI. CONCLUSIONS

We have explored the problem of thyroid nodule classification. The overall process includes preprocessing of ultrasound images, image data augmentation and classification by transfer learning. The preprocessing step focuses on the extraction of ROI. For the data augmentation part, we have tried two different means, the traditional methods that augments data by direct changing the original images, and the convolutional method that uses a small three-layer convolutional network to generate new images. The former one creates new images at an order of magnitudes of N from N initial images, while the convolutional network is capable of generating at an order of magnitudes of N^2. As for the final classification step, a pre-trained residual net is adopted for transfer learning, and three sets of experiments are conducted based on different datasets, and the performances are evaluated by three indexes, the accuracy, sensitivity and specificity. The final results have shown the effectiveness of our proposed convolutional network in generating new thyroid ultrasound images. The best accuracy attains 93.75%, which is slightly higher than Experiment 1 and surpasses the classification accuracy of other methods proposed by [6], [7] and [8]. Consequently, the experiment has proved on the one hand the effectiveness of transfer learning in classifying ultrasound images of thyroid nodules, on the other hand, it has also revealed the effectiveness of our proposed augmentation net in image data augmentation.

In future work, we plan to apply a generative adversarial network (GAN) to realize a more universal method for general medical image data augmentation and thus to pursue a more detailed classification in various kinds of diseases.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China [grant numbers 61473192, U1401240, 61075086]; and the National Basic Research Program of China [grant number 2014CB046302]; and the Project of STCSM [grant number 17441901000].

REFERENCES

[1] Suzuki, Kenji. "Overview of deep learning in medical imaging." Radiological Physics & Technology 7(2017):1-17.

[2] Moon, W. J., et al. "Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study." Radiology247.3 (2008):762.

[3] Ginneken, Bram Van. "Computer-aided diagnosis in chest radiography." IEEE Transactions on Medical Imaging 20.12(2001): 1228.

[4] Zakeri, F. S., H. Behnam, and N. Ahmadinejad. "Classification of benign and malignant breast masses based on shape and texture features in sonography images. " Journal of Medical Systems 36.3 (2012):1621-1627.

[5] Li, L. N., et al. "A computer aided diagnosis system for thyroid disease using extreme learning machine. " Journal of Medical Systems36.5(2012):3327-3337.

[6] Liu, Tianjiao, et al. "Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features." IEEE International Conference on Acoustics, Speech and Signal ProcessingIEEE, 2017.

[7] Luo, S., et al. "Thyroid nodule classification using ultrasound elastography via linear discriminant analysis. " Ultrasonics 51.4(2011):425-31.

[8] Legakis, Ioannis, et al. "Computer-Based Nodule Malignancy Risk Assessment in Thyroid Ultrasound Images." International Journal of Computers & Applications 33.1(2011):29-35.

[9] Ding, Jianrui, et al. "Multiple-instance learning with global and local features for thyroid ultrasound image classification." International Conference on Biomedical Engineering and Informatics IEEE, 2015:66-70.

[10] Liu, Bo, et al. "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images." Pattern Recognition 43.1(2010):280-298.

[11] J. Wang., “The Effectiveness of Data Augmentation in Image Classification using Deep Learning.”

[12] Pedraza, Lina, et al. "An open access thyroid ultrasound image database." International Symposium on Medical Information Processing and Analysis 2015:92870W-92870W-6.

[13] D. Gaitini, R. M. Evans, and G. Ivanac, Chapter16: thyroid ultrasound. EFSUMB Course Book, 2011.

[14] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." Computer Vision and Pattern Recognition IEEE, 2016:770-778.